

# Exhibit A

UNITED STATES DISTRICT COURT  
DISTRICT OF NEW JERSEY

IN RE: VALSARTAN N-  
NITROSODIMETHYLAMINE (NDMA)  
CONTAMINATION PRODUCTS  
LIABILITY LITIGATION

HON. ROBERT B. KLUGER

Civil No. 19-2875 (RBK/JS)

**DECLARATION OF JONATHAN JAFFE**

I, Jonathan Jaffe, declare, under penalty of perjury, pursuant to 28 U.S.C. § 1746,  
as follows:

I, Jonathan Jaffe, am the founder and owner of Its-Your-Internet, a technology and ESI  
discovery consulting firm, established in 2008.

As the Court is aware, I have assisted Plaintiffs in this litigation on ESI discovery issues,  
including more recently the responses to the proposal of TEVA to utilize what they have  
termed “CMML” or what is referred to in the industry as Continuous Active Learning “CAL”  
or “TAR 2.0” in their review process.

I was part of the July 8<sup>th</sup> telephonic meet and confer with TEVA and their ESI experts. I  
attended by telephone the hearing held by the Court on July 15<sup>th</sup>.

I have also been privy to all the communications via letter and e-mail between the parties as  
well as submissions to the Court regarding this matter.

1 For the past fifteen years, I have provided ESI discovery and litigation consulting for  
2 attorneys in complex litigation throughout the lifecycle of each litigation (from inception to  
3 the ultimate resolution). In multiple multi-district litigations inclusive of this one, I have  
4 assisted with initial efforts by the parties to reach agreement on search methodologies, as  
5 well as to discover the structure of Defendants' systems for maintaining ESI and physical  
6 materials. In that capacity I worked with Plaintiffs' counsel in the In Re Benicar  
7 (Olmesartan) Products Liability Litigation that was previously litigated before this Court.

8  
9 I graduated magna cum laude from Columbia University with a BA in Economics-  
10 Mathematics in 1999.

11  
12 For the Court's reference, a copy of my *curriculum vitae* is attached hereto as **Exhibit A**.

13  
14 While I have been retained by the Plaintiffs, the opinions I express below are independent of  
15 any party.

16  
17 **ALLOWING TEVA TO USE TAR/CMML/CAL TO PRIORITIZE AND FILTER**  
18 **DOCUMENTS FOR PRODUCTION IS HIGHLY PREJUDICIAL TO PLAINTIFFS**

19  
20 At the hearing on July 15<sup>th</sup>, and in submissions to the Court, TEVA has made multiple  
21 representations as to the nature of CAL, specifically as to what it can do and what it cannot.  
22 As set forth in more detail below, there are fundamental flaws in TEVA's approach,  
23 including for example fundamentally incorrect assumptions underlying the arguments as to  
24 the value of this system for this particular review for production, the failure to faithfully work  
25 to satisfy the Plaintiffs' prioritization requests, the proposed application of the CAL system  
26 to the truncated document set already narrowed by the search terms, rather than to the entire  
27 document set as is intended, TEVA's refusal to engage in a transparent discussion and  
28 collaboration designed to establish the reasonable tagging decisions to apply at the outset,  
29 and TEVA's refusal to involve the Plaintiffs throughout the process to ensure faithful  
30 application of the tags and continuous quality control and adjustments to continually guide  
31 the learning process of the system.

**IMPROPER APPLICATION OF CAL AFTER SEARCH TERMS**

In the search term negotiation process in which I was heavily involved, Plaintiffs negotiated with the understanding that TAR and CAL would not be used in the review and production process, as no Defendant indicated that it would be using TAR or CAL for review and production.

If a CAL or TAR process is used, it should be applied across the entire set of collected documents. That is how the systems, including the system at issue here, are best utilized. See: “White Papers” attached to the Plaintiffs’ July 14, 2020 submission.

Plaintiffs agreed to search term modifications that were only reasonable if TAR and CAL were not going to be used. If this intent had been disclosed, my strong advice would have been not to agree to the modifications and that the present objections would have been presented at that time.

**TEVA’S ARGUMENT IS BASED ON A FALSE PREMISE**

Notably, in their late evening submission on July 14<sup>th</sup> to the Court, the TEVA Defendants wrote the following in footnote #1:

CMML is a technology platform that utilizes a standardized, defensible process to analyze textual content to allow for rapid relevant document discovery. The predictive scores from the CMML model are used to prioritize document review, ensuring that the most relevant documents are reviewed (and produced) first.

Defense counsel repeated this assertion that the prioritization afforded by CMML would result in Plaintiffs receiving the **most relevant documents** in a prioritized manner (though never specifically explaining how each request for prioritization would be fulfilled), during the July 15<sup>th</sup>, 2020 conference with the Court.

1 These statements are not correct. Counsel and TEVA's statements are premised on an  
2 inaccurate assumption as to a link between the relevancy scores and the *degree* of relevance.

3  
4 The predictive scores (usually expressed in a range between zero and one) outputted by a  
5 CAL system are a representation of how well matched the machine believes a particular  
6 document is to the set of documents previously tagged.

7  
8 The TEVA Defendants have represented that they are only using a single tag. That is  
9 Responsive or Not (I often herein refer also to relevant/not relevant, as TEVA has at  
10 sometimes referenced relevance rather than responsiveness, and since that is the touchstone  
11 of this process).

12  
13 They are not indicating to the system the weight of the relevance nor are they indicating the  
14 topic of relevance.

15  
16 Accordingly, a perfect score of 1.0 **only** will indicate that the machine finds the document to  
17 be in high correlation with other documents **currently** marked as relevant.

18  
19 It does **NOT** bear any indication on whether the document is actually relevant, on which  
20 topic it is relevant, nor its degree of relevance with respect to that topic.

21  
22 A document with a perfect score of 1.0 may only be **marginally** relevant to the litigation, but  
23 well matched to many prior reviewed, but marginally relevant documents.

24  
25 Conversely, a document with a score of 0.0002 might be **extraordinarily relevant** to the  
26 litigation, but just different or new in comparison to the rest of the documents that were  
27 marked relevant, or it may be very similar to documents previously marked as not relevant.

28  
29 Let's take three concrete examples.

30  
31 (1) A testing report showing NDMA contamination

1 (2) An email addressing NDMA contamination

2 (3) A chromatogram showing NDMA contamination

3  
4 A testing report showing NDMA contamination is not merely relevant, but substantially  
5 relevant to the issues in the litigation. However, suppose that reviewers have already marked  
6 1000 emails with attached testing reports that show test results across many disparate items  
7 as not relevant. The system has effectively been “trained” that emails transmitting testing  
8 reports are generally not relevant. So, when the system examines this next testing report,  
9 where there’s a singular notation of NDMA in the listed chemicals found, it is not  
10 unreasonable to expect that the system, even a very intelligent system, will assign a  
11 predictive score close to zero. Thus, because irrelevant testing reports are common, an email  
12 referencing a highly relevant report may get a low score, let’s say of 0.0002. I set forth a non-  
13 zero score because there is some textual difference from prior marked documents. This  
14 means this highly relevant report **AND THOSE REPORTS LIKE THIS ONE** will be  
15 deprioritized **even if there are multiple reports similar to this one, but the system was**  
16 **just trained that most reports are irrelevant.**

17  
18 Let’s move on to my second example, an early email on NDMA contamination. Let’s say  
19 that reviewers have reviewed many emails, marking many relevant for many different  
20 reasons, but the next email is the very first email on NDMA contamination. The machine’s  
21 algorithm will naturally not recognize a clear pattern between this email and the previous  
22 ones marked relevant since it has been “trained” on the first emails marked as relevant, which  
23 are numerous at this point. So, this email on NDMA contamination will naturally be assigned  
24 a lower predictive score closer to zero than one since it may not be a good “match”  
25 linguistically to what was previously marked. The score assigned here does not accurately  
26 indicate the extent to which this email is relevant. Let’s suppose the machine gives it a score  
27 of 0.25. **BUT worse yet the bias persists.** Let’s suppose even though this email has a lower  
28 score, it is randomly inserted into the next review queue and reviewed. Let’s even suppose  
29 the reviewer marks this email as relevant. Even if the reviewer marks this email as relevant,  
30 since this is only 1 email out of let’s suppose 10,000 marked relevant, **the next email on**  
31 **NDMA contamination will still get a score closer to zero than one.** In fact, there is no

1 guarantee that the next email on NDMA contamination will even get a higher score than the  
2 first. Rare items are thus biased in their predictive scores since even the most intelligent  
3 system can only base its predictions using previous patterns. But clearly the most relevant  
4 items are by definition going to be relatively fewer in number.

5  
6 TEVA included a footnote that the CMML system is scoring based upon the analysis of the  
7 **textual content of a document**. Even if that system worked perfectly, there are multiple  
8 categories of documents where the textual content is imperfect or wholly absent. Documents  
9 where there is an image / graphic that is highly relevant, hard copy documents that were  
10 OCR'd, audio and video files, and many more could have been picked up by the search terms  
11 because of their original file name, folder path they were in, or a portion that did get properly  
12 OCR'd, yet given very low predictive scores by the CAL software.

13  
14 Let's take my third example, a chromatogram attached to an email that shows an NDMA  
15 spike. Chromatograms are expressed as images. As such they are effectively invisible to  
16 TEVA's vendor's selected CAL software.

17  
18 Again, the scoring has no correlation to how relevant a document is. The best that can be said  
19 is that the CMML system will show reviewers sets of documents more likely to be similar to  
20 documents they have previously indicated were relevant. It is a misread to impute to the  
21 score any meaning other than whether the machine is saying a document matches the existing  
22 marked relevant set.

23  
24 Unfortunately, reviewers who see scores may be biased by them. A reviewer who sees the  
25 computer's opinion of relevancy is 0.0002 is only human. This bias becomes self-reinforcing  
26 since if reviewers see a consecutive series of low scored documents that they agree are  
27 irrelevant, they will begin to "trust" the computer's judgment. The weight of marking rarer  
28 documents as irrelevant has a much stronger adverse effect than the positive effect of  
29 marking a rare document relevant in a CAL review. As the corpus of relevant documents  
30 increase, when a highly relevant document is mistakenly marked as irrelevant, that "trains"  
31 the system that that variation is likely irrelevant.

1  
2 So, counsel's and TEVA's representation that their CMML / CAL process will automatically  
3 prioritize the production of **the most relevant** documents does not stand the weight of even a  
4 cursory examination.

5  
6 Clearly, the only prioritization is of responsive documents in the already narrowed set of  
7 documents selected by use of the narrowed search terms, and based solely on the definition  
8 of responsiveness that TEVA refused to disclose during the meet and confer call, and in my  
9 opinion there is no reasonable basis to assert that the prioritized documents are **the most**  
10 **relevant**. In fact, as set forth above, some of the most relevant documents are unlikely to be  
11 prioritized (and due to bias may even potentially be excluded) than if the ordered linear  
12 review, with targeted prioritization, is implemented. Moreover, since search terms were  
13 already used to substantially narrow the document set, some quantity of documents that were  
14 excluded would have been captured by the CAL process if it were applied to the entire  
15 document set.

16  
17  
18 **ANY SYSTEM IS ONLY AS GOOD AS ITS INPUTS**

19  
20 In some ways, these complex CAL systems can be thought of in very simple terms, as akin to  
21 a calculator.

22  
23 In elementary school, when one is first introduced to a calculator, it appears to be a magical  
24 tool. For any simple numerical problem, you just enter in the values and operations, and *voila*  
25 out appears an answer.

26  
27 For the slightly more complex operations, such as square roots, it's even more magical that it  
28 can provide an answer so quickly.

29  
30 Unfortunately, one of the first hard lessons you learn with a calculator is that merely getting  
31 an answer does not mean you got the **right** answer.



1  
2 At the July 15, 2020 hearing, defense counsel talked at length and at multiple points about  
3 the intelligence of TEVA's machine.

4  
5 TEVA's machine is like any other. It is only as good as its inputs. It may be a very  
6 sophisticated calculator, but Plaintiffs cannot judge or, more important, ensure its precision  
7 and accuracy without examining the inputs along with the outputs. Moreover, as stated  
8 herein, the CAL system should be applied to the entire collected document set for production  
9 purposes, not *ex post facto* on the narrowed set created by prior application of search terms.

10  
11  
12 **CAL SYSTEMS AT THEIR HEART ARE PATTERN RECOGNIZERS AND**  
13 **CODIFIERS OF BIAS**  
14

15 These CAL systems are much more sophisticated than a calculator, but conceptually and  
16 practically, they can be thought of as sophisticated pattern recognizers.

17  
18 There is another way to think of CAL systems. They codify bias.

19  
20 I do not mean to impute any negative connotation to the word bias. What I mean to say is that  
21 the bias of early reviewers naturally becomes self-reinforcing in any CAL system.

22  
23 The conceit of CAL is that it attempts to standardize those individual biases across the whole  
24 set of reviewers.

25  
26 With CAL, unlike in a linear review, decisions made by early reviewers strongly influence  
27 which documents are shown next. Even if scores are not shown to reviewers, the very fact  
28 that a reviewer gets series of mostly relevant or mostly irrelevant documents in effect "trains"  
29 the reviewer. An irrelevant document in a sea of relevant documents is more likely to be  
30 marked as relevant. Conversely, and of greater concern as to bias against the Plaintiffs,

1 relevant documents, even highly relevant, are less likely to be detected as relevant by a  
2 reviewer if they are presented in a sea of documents deemed irrelevant.

3  
4 Ultimately, these observations simply demonstrate what is apparent. Without the full  
5 engagement of the Plaintiffs from the outset, which would have been a time consuming  
6 process, and would require ongoing engagement on a going forward basis in order to combat  
7 and quality check the biases, the production resulting from application of CAL to the  
8 truncated set of documents utilized provides no reasonable assurance of fairness or accuracy.  
9

### 10 11 A HAY STACK WITH MANY NEEDLES

12  
13 Review is not uncommonly compared to as finding the needles in the haystack. The hay and  
14 needles represent the documents collected. The needles are the documents relevant to the  
15 litigation.  
16

17 But as discussed before, not every needle is the same. Let's say that the sharpest needles  
18 represent the documents **most relevant** to the litigation whereas blunt needles are relevant,  
19 but not necessarily the most relevant.  
20

21 Our search term process, through the multiple rounds, was designed to weed out hay. Of  
22 course, in reducing the size of the pile, we accepted that we probably left out some needles,  
23 but we agreed that there was likely a greater concentration of needles in what we were  
24 retaining for review.  
25

26 Since we do not know otherwise and as we select batches to examine at random, we have to  
27 assume that the needles are evenly distributed in the hay.  
28

29 In a linear review, a random bunch of hay is grabbed and examined. The likelihood of that  
30 hay containing needles both blunt and sharp is equally likely for any other batch. So, while  
31 there is hay to sort through, there is no prejudice. We all accept that in reviews of large

1 volumes of documents with limited time, some needles might be missed. But the important  
2 part is that there is no inherent bias in the process. And, the search terms were applied to the  
3 entire stack of hay.

4  
5 CAL is very different, especially as proposed by TEVA to be only be applied to the hay and  
6 needles that were identified by the search terms – excluding a vast expanse of hay and by  
7 extension, needles, that should have been searched from the outset.

8  
9 In a CAL simulated review process, the first few batches of hay are random. Reviewers pull  
10 out the needles they see in these batches.

11  
12 BUT since blunt needles are much more common, statistically, there will be a lot of blunt  
13 needles pulled and perhaps a few sharp needles.

14  
15 The problem compounds in the subsequent batches.

16  
17 As the system is “trained,” it recognizes that the blunt needles are needles that are being  
18 identified and pulled out of the hay. So, the next batches are weighted to pull more blunt  
19 needles and less other, the other being sharp needles and hay.

20  
21 After some time, all the blunt needles have been pulled out of the stack. Now, when new  
22 batches are examined, the system preferences hay that looks like blunt needles, not  
23 necessarily the sharp needles because there is likely more hay that resembles needles in the  
24 remaining stack than actual needles. So, reviewers start complaining about the waste of time  
25 going through all of this hay that’s just hay. They may find some needles, but not think it  
26 worth the effort.

27  
28 The problem is that the sharpest needles are likely, by simple statistical inference, to remain  
29 in the stack. Pulling samples from the remaining stack will just validate the review that all of  
30 the blunt needles have been produced.

1 Plaintiffs and Defendants have different perspectives. Plaintiffs believe sharp needles exist  
2 and are often rare. Defendants believe they do not exist. The purpose of the review is to  
3 identify for production the documents that should be produced.  
4

5 Plaintiffs and Defendants have already worked to refine the search terms, in an effort to  
6 identify the blunt and sharp needles, and the narrowed the search terms extensively, and  
7 made other reasonable concessions to efficiency. For example, when words were showing  
8 up in email footers, we worked collaboratively to exclude those hits. When inspectors' names  
9 were common, *e.g.* Patel, we worked to exclude those hits. Should Defendants during their  
10 review approach Plaintiffs on a category of documents such as company newsletters that are  
11 challenging, Plaintiffs have already told Defendants they would work with Defendants to  
12 fashion a mutually agreeable solution. The proposed application of CAL to the already  
13 heavily narrowed set of documents was never contemplated, and if it was known that this  
14 methodology would be used there would not have been a limitation of the part of the  
15 haystack to be searched – the system is built to search the entire hay stack.  
16

17 If efficiency and saving time and money is truly the goal, using CAL to identify voluminous  
18 sets of documents least likely to be relevant, for example, news clipping services that may  
19 have gotten unintentionally captured by search terms and could be narrowed through  
20 additional searches before reviewing them all, would be more effective. As in the classical  
21 solution to burning down the haystack to find the needles as opposed to building ever more  
22 complicated needle finding machines, my point is that there are other ways of significantly  
23 cutting down the review burden that are less likely to introduce prejudice.  
24  
25  
26

### 27 **MORE COMPLEX THAN A HAYSTACK**

28  
29 In complex litigation, we do not simply have hay, blunt and sharp needles. Not only are  
30 documents more or less relevant, but they are more or less relevant with respect to a large

1 number of topics. For example, in this litigation, a document could be more or less relevant  
2 to what a company told regulators versus what they knew internally.

3  
4 Instead of a haystack, picture a forest of documents. Surveying the forest, we want to  
5 distinguish all of the wildlife from the trees. It is too simplistic to just categorize into tree or  
6 not. We want to categorize into birds, insects, larger creatures, and unexpected visitors.

7  
8 Unfortunately, if we configured a CAL system in the same manner as TEVA's proposal, it  
9 would tend to get "stuck" on a particular topic until it ran out of that topic. If sparrows are  
10 common in our forest, reviewers would be shown pictures of every sparrow until they were  
11 exhausted. Then the CAL system would get "stuck" on the next random relevant category  
12 that grabbed its attention. It would become fixated and refuse to show pictures of anything  
13 else.

14  
15 Why? Because in TEVA's configuration, the system has no inputs on topic, no assurance of  
16 contextual diversity, and no feedback on degree of relevance. It can only codify bias based  
17 upon what it is told.

18  
19 The natural consequence is an inverted production where the least interesting relevant  
20 documents are produced first and in great quantity. In our forest survey, we want to learn  
21 about the rarer migratory animals in each specific category, not just get a dump of pictures of  
22 ants, sparrows, and squirrels.

23  
24 In our more complex haystack, an early review cutoff deprives the receiving party of the  
25 most important information across multiple different topics.

26  
27 In this litigation, there are numerous key topics fundamentally distinct from one another,  
28 represented by different and overlapping sets of documents. A number of these topics have  
29 yet to be identified and will only become apparent during document reviews and depositions.  
30  
31

## EXAMPLES OF UNEXPECTED BIAS IN REAL WORLD AI SYSTEMS

Although machine learning systems (AI) are expected to be neutral, fair, and intelligent, there are plenty of ways that bias can be unwittingly introduced, negating the perceived impartiality of the output.

As Karen Hoa wrote in her review of AI Bias in MIT Technology Review last February, <https://www.technologyreview.com/2019/02/04/137602/this-is-how-ai-bias-really-happens-and-why-its-so-hard-to-fix/>, bias can be introduced during framing the problem, collecting the data, or preparing the data. It can be difficult to detect:

“The introduction of bias isn’t always obvious during a model’s construction because you may not realize the downstream impacts of your data and choices until much later. Once you do, it’s hard to retroactively identify where that bias came from and then figure out how to get rid of it. In Amazon’s case, when the engineers initially discovered that its tool was penalizing female candidates, they reprogrammed it to ignore explicitly gendered words like “women’s.” They soon discovered that the revised system was still picking up on [implicitly gendered words](#)—verbs that were highly correlated with men over women, such as “executed” and “captured”—and using that to make its decisions.”

When Amazon used AI in their hiring, they discovered afterwards it was biased against women. When NVIDIA developed an AI to unblur faces, they discovered the system was outputting faces that transformed the faces of people of color into faces that looked distinctly white, <https://www.theverge.com/21298762/face-depixelizer-ai-machine-learning-tool-pulse-stylegan-obama-bias>. Similarly unexpected bias has been documented even with large sets of training data, such as the 14MM images tagged as part of the ImageNet data set where matching on neutral terms such as “programmer” yielded images of white men: <https://www.wired.com/story/ai-biased-how-scientists-trying-fix/>. Facial recognition algorithms routinely struggle with bias: <https://www.wired.com/story/best-algorithms-struggle-recognize-black-faces-equally/>. Even Google has struggled to eliminate contextual bias in simple search prompts: <https://www.wired.com/story/google-autocomplete-vile-suggestions/>. This has also impacted the Courts as the well documented bias in the COMPAS

1 system used to predict rates of recidivism: [https://www.propublica.org/article/how-we-](https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm)  
2 [analyzed-the-compas-recidivism-algorithm](https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm).

3  
4 Accordingly and unsurprisingly, in August 2019, the American Bar Association, recognizing  
5 the challenges of bias in AI system passed Resolution 112:

6 ([https://www.americanbar.org/content/dam/aba/images/news/2019/08/am-hod-](https://www.americanbar.org/content/dam/aba/images/news/2019/08/am-hod-resolutions/112.pdf)  
7 [resolutions/112.pdf](https://www.americanbar.org/content/dam/aba/images/news/2019/08/am-hod-resolutions/112.pdf))

8  
9 RESOLVED, That the American Bar Association urges courts and lawyers to address the  
10 emerging ethical and legal issues related to the usage of artificial intelligence (“AI”) in the  
11 practice of law including: (1) bias, explainability, and transparency of automated decisions  
12 made by AI; (2) ethical and beneficial usage of AI; and (3) controls and oversight of AI and  
13 the vendors that provide AI.

14  
15 The resolution which was adopted at the Annual Meeting and accompanying announcement,  
16 [https://www.americanbar.org/groups/business\\_law/publications/committee\\_newsletters/legal](https://www.americanbar.org/groups/business_law/publications/committee_newsletters/legal_analytics/2019/201908/ai_law/)  
17 [\\_analytics/2019/201908/ai\\_law/](https://www.americanbar.org/groups/business_law/publications/committee_newsletters/legal_analytics/2019/201908/ai_law/), together make clear that any adoption or use of TAR/CAL  
18 should be done transparently with an eye on bias.

19  
20  
21 **CAL PROPERLY DONE**

22  
23 As with any tool, there are proper and improper applications of CAL systems. As a  
24 threshold, as set forth above, the system is designed to be applied to the entire document set.  
25 In an effort to address a fair application of this system, and to respond to the Court’s inquiry  
26 regarding necessary tagging, it is my opinion that in a proper application of CAL in this or  
27 other similar complex litigations, the producing party must:

28  
29 (1) use multiple tags / vectors for relevance as negotiated with the receiving party instead of a  
30 universal single relevant/not relevant tag, i.e. relevant because... In the absence of a  
31 transparent meet and confirm with TEVA’s counsel and vendor, it is difficult at this point to  
32 identify with precision all of the additional tags that should be used. Having said that, at a

1 minimum, in this litigation, high level concepts such as contamination, chemistry issues,  
2 quality control, quality assurance, sales/financial, marketing, cGMP, adverse events,  
3 inspections, regulatory, chromatography and mass spectrometry testing are just some of the  
4 topics that would need to be coded in order for a CAL system to be able to properly prioritize  
5 and mark for relevance.

6  
7 (2) use at least one tag to indicate the relative weight of relevance, e.g. clear, probable,  
8 possible, neutral, contextual  
9

10 (3) implement prioritization through targeted initial seed sets  
11

12 a) run searches for the most relevant documents; this has to be done categorically for each  
13 category

14 b) mark the relevant documents in each category (using senior reviewers)

15 c) feed back into the system the relevant documents (excluding ones marked irrelevant) to  
16 train it in each categorization

17 d) use senior reviewers to conduct quality assurance

18 e) share initial seed sets with the receiving party

19 f) accept initial seed sets from the receiving party  
20

21 (4) use random sampling and CAL at continuous points throughout the process to re-examine  
22 documents reviewers marked non-relevant  
23

24 a) reintroduce any documents marked as not relevant in error back into the CAL system

25 b) rerun CAL to find similar documents marked as not relevant that should have been marked  
26 as relevant  
27

28 (5) do training side by side with the receiving party until the receiving party is in agreement  
29 on what constitutes relevant documents  
30



(6) on a separate CAL project, use tagging to weed out document types (ex. vacation requests, janitorial supply invoices, etc.) the producing party believes are categorically not relevant (but these would need to be logged so that if the receiving party disagreed that these documents are not relevant there is transparency and that decision could be challenged)

(7) ensure the introduction of contextual diversity in subsequent review batches; this can be set through appropriate configuration flags

(8) suppress any CAL relevance scoring when reviewers are reviewing a document to avoid introducing bias

(9) set a threshold score above which documents will be automatically produced without any further review where CAL has scored those documents with a mutually accepted high degree of probable relevancy in one or more categories

(10) after reviewing an agreed upon percentage of documents not less than 10%, automatically produce (subject to privilege review) all documents the CAL system has marked as relevant above a certain threshold point

(11) not automatically apply CAL training to subsequent custodians or sets of documents that are expected to be significantly different from the prior reviewed corpus

(12) determine *a priori* whether the set of collected documents is large enough and diverse enough to support CAL or otherwise inappropriate for CAL; not every set of documents would be appropriately reviewed by CAL

(13) abstain from using CAL for certain categories of documents where relevance is near impossible to recognize outside of context or where CAL is ill suited to do a proper linguistic analysis, ex: very short emails, spreadsheets, chromatographs, other images, audio files, video files, OCR'd files

(14) conduct extensive targeted sampling (with full disclosure of all sample documents to the receiving party) before proposing a discard point if the producing party is not using CAL to only prioritize the review (in the situation here, where the initial groundwork has not been done, all documents targeted for exclusion should be manually reviewed).

**UNWORKABLE AT THIS POINT IN THE PROCESS**

During the July 15<sup>th</sup>, 2000 hearing with the Court, Teva's counsel cited to the TAR/CAL protocol that was adopted In Re Broiler Chicken Antitrust Litigation, Case No. 1:16-cv-08637 N.D. Illinois.

While a discussion of CAL/CMML at the outset would have given the Parties enough time to properly implement joint reviews and an overall protocol for review, at this point in the process with the production of documents underway, establishing a workable CAL process that is fair and equitable and that will result in timely, fully prioritized productions, is not reasonably achievable, and the effort would impose a significant distraction on the Plaintiffs. The sudden suggestion by ZHP and Mylan that they might elect to use CAL only would multiply the difficulty.

The protocol in the Broiler Chicken Antitrust litigation was set out right at the beginning of the process, not foisted on at the last minute. The TAR/CAL system to be used had to be disclosed up front, prior to any search terms being negotiated. The Plaintiffs could opt out custodians from the search terms and the TAR/CAL process and subject them solely to the TAR/CAL process. Those threshold steps did not occur here, and cannot occur here, due to the belated revelation of Teva's intent to do this. Unfortunately, the validation process laid out in that protocol was very limited and inadequate for a litigation of this size and complexity. A great deal of work would need to be focused on establishing an appropriate validation process for this litigation.

1 The TEVA Defendants are pioneering CAL review systems within this MDL. They intend to  
2 do so unilaterally. They likely intend to use these systems to support future arguments for  
3 thresholds of scores below which they will not review documents or will offer cursory  
4 examination.

5  
6 Contrary to any assertion that the system is currently “learning” intelligently, the best  
7 evidence is that the system is preferencing more numerous documents over documents that  
8 are rarer or different, but no less relevant or critical to proving this case.

9  
10 This is not a matter of sitting down once with TEVA to work out a protocol. Due to the  
11 nature of CAL, Plaintiffs would have to be involved in the actual responsiveness review.  
12 Plaintiffs would have to be intimately involved in the training of the system. That is not  
13 possible at this point.

14  
15 It is clear that Defendants are beta testing this system. This is a complicated system with  
16 complex side effects. Errors and missteps are costly both in time and money to fix.

17  
18 At this point, doing CAL properly would add a potentially drastic amount of time to the  
19 review and require collaborative retraining of the existing CAL system. Again, this would be  
20 a significant distraction where the Plaintiffs are now ingesting document productions on a  
21 rolling basis from numerous defendants.

### 22 23 24 **INFEASIBILITY OF VALIDATING POST PRODUCTION**

25  
26 It is infeasible for TEVA to ask the Plaintiffs meaningfully to evaluate a CAL process post  
27 the initial productions.

28  
29 First, the initial productions will, by definition, contain relevant documents. How are  
30 Plaintiffs to determine which categories of relevant or prioritized documents are missing,

1 especially if the Defendants simply argue that they do not exist? There is no way for  
2 Plaintiffs to know.

3  
4  
5  
6 **PLAINTIFFS ARE UNDULY PREJUDICED EVEN IF EVERY DOCUMENT IS**  
7 **ULTIMATELY REVIEWED**  
8


9 For the reasons outlined before, rarer but highly relevant documents, “diamonds,” will likely  
10 be pushed to the last of the rolling productions. That effectively negates the intent of a rolling  
11 production, and absolutely negates the prioritization by Plaintiffs. This will result in  
12 Plaintiffs having to search a potentially large production to identify documents that in many  
13 cases should have been produced at the very outset.  
14  
15

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14

**AFFIRMATION**

Pursuant to 28 U.S.C. § 1746, I declare under penalty of perjury of the laws of the United States of America that the foregoing is true and correct.

Executed on this 24<sup>th</sup> day of July, 2020, at Forest Hills, NY.



Digitally signed by Jonathan Jaffe  
DN: cn=Jonathan Jaffe,  
o=www.its-your-internet.com,  
ou, email=jjaffe@its-your-  
internet.com, c=US  
Date: 2020.07.24 13:59:29  
-04'00'

---

Jonathan Jaffe

# Exhibit A

**JONATHAN JAFFE**

www.its-your-internet.com

Forest Hills, New York 11375

(866) 526-1836

[jjaffe@its-your-internet.com](mailto:jjaffe@its-your-internet.com)[linkedin.com/in/jonathankjaffe](https://www.linkedin.com/in/jonathankjaffe)**EXPERT TECHNOLOGY + LITIGATION ESI CONSULTING****ESI Technical Representative...Federal Expert Witness...ESI Discovery Consulting  
Database/Data Analysis...Statistical Sampling...Gap Analysis...ESI Forensics****Its-Your-Internet****2008-present**

Its-Your-Internet is a bespoke general software and ESI litigation support consultancy. It was established in March 2008.

***Founding Owner, 2008-present;***

- Plaintiffs' Designated ESI Representative for Over 500 Firms on Multiple Mass Tort Litigations involving Product Liability with Medical Devices and Drugs (hundreds of thousands of individual plaintiffs collectively) extensive engagement from early discovery through trial
- Plaintiffs' Designated ESI Representative in Multiple Class Action Suits (many involving Fair Credit Reporting Act violations in consumer reporting)
- Plaintiffs' Designated ESI Representative in Attorney General Suits involving Product Liability
- Expert Witness in Federal Cases on ESI Issues
- Database Analysis (Consumer Reports, Adverse Event/Safety, Clinical Trial, Registry, Call Note, Sales and Marketing, Statistical Analysis)
- Assisting with Class List development and certification
- Working with Extremely Large Document Productions (100MM+ pages)
- Working with Extremely Large Database Productions and Extracts
- Analysis of Statistical Results from Clinical Studies
- Trial Support (exhibit list document selection and preparation, helping find documents used with adverse witnesses and in cross examination, translating finds into usable exhibits, demonstratives, jury charges)
- Plaintiff Document Production (Workflow, Redaction, Privilege Review, Staging and Producing)
- Built Bespoke Production Review Software
- Discovery Format and Production Negotiation
- Redaction / Privilege Review Management
- Keyword Search Term Recommendations, Predicative Coding Recommendations
- Custodial Production Forensics
- Custodial Production Storage Recommendations
- Drafting ESI Related Interrogatories, Requests for Production, Requests for Admission
- Deduplication, Gap Analysis, Statistical Sampling, Forensic Analysis, Discovery Issues Analysis
- Assisting Attorneys on 30(b)(6) Depositions, Meet and Confers, Hearings, Motions, Draft Orders, Stipulations, Rule 34 issues
- Supporting other Expert Witnesses
- Creating Supporting ESI Reports for Court Hearings
- Deposition Workup and Providing Deposition Assistance

- Supervising and participating (design, coding, analysis) over a large variety of projects, most with significant teams, inclusive of:
  - Litigation Support Custom Applications / Software / Data Analysis
  - Subscription Site for M&A Information, OAuth2 Authentication, Graph Analysis, and Internal Application Development for Investment Bank
  - Real-time Chat Extension for Large VoIP Provider
  - Redesign of Class Action Support System for a Top 3 Legal Administrative Services Firm (Billions of Dollars of Administration)
  - LMS and Exam SaaS applications for West African Market
  - Mobile West Africa #1 App – FindAMed
  - Subscription Based Site for Theater and Performing Arts Companies
  - School Based Management Systems Development
  - Raffle Contest Site and Variations
  - Writing and Support for Core Remake of Multi-Million Page Web Site
  - SEO and SEM Management for Multiple Clients
  - Finalist in Multi-Billion \$ RFP for NYC DOE
  - > 50 original web sites inclusive of e-Commerce
  - Countless consultations with Entrepreneurs
- Multiple Entrepreneurial Projects Using Bleeding Edge Technology
- Established a Dozen Key Reciprocal Partnerships
- > 30 Talks and Lectures on Technology and Social Media for Chambers of Commerce, SBA, Charities and Non-Profit, Business Groups and Organizations
- Day to Day Resource Management, Staffing, Administration
- Oversight of Remote and Offshore Resources
- Branding, Marketing, Materials (for Firm and Clients)

General Technologies: NoSQL, SQL, C#, PHP, NodeJS, Python, Django, NPM, PaaS, Heroku, HTML5, BackboneJS, EmberJS, AngularJS, MongoDB, Cloudant, Redis, MSSQL, Oracle, MySQL, MariaDB, Postgres SQL, ExpressJS, Git/GitHub, GruntJS, E2E Testing, ASP.NET MVC, ASP.NET, Cordova PhoneGap, Amazon Web Services (S3, CloudFront, EC2), Azure, XML, XML Schema, LESS, CSS3, SocketJS, Bower, Joomla, Drupal, WordPress, Office365, SharePoint, Salesforce, JIRA, Photoshop, Illustrator, GIMP, RESTful and Agile Methodologies, Quantum Computing (early exploration).

Legal Review Platforms: Relativity, Catalyst Insight, Everlaw, iConect-Xera, Eclipse, Disco, Crivella

Forensic Software: FTK Imager, Encase, Cellebrite PA UFED, Transend Migrator, Magnet IEF

#### **Weitz & Luxenberg, P.C.**

**2002-2008**

Weitz & Luxenberg is one of the nation's original and foremost firms in mass tort asbestos litigation handling nearly one million simultaneous pending court actions, 250,000 bankruptcy claims, and over 50 other personal injury litigations.



***Manager of Software Development, 2003-2008******Sr. Project Manager, 2002-2003***

- Federal Expert Witness resulting in substantial eDiscovery Sanctions
- Collection of documents for litigation; Validation; Forensic Analysis
- Coordinated activities of 5 teams, comprising 20 resources
- Achievements included: establishing and overseeing the 4<sup>th</sup> largest personal injury law firm site, [www.weitzlux.com](http://www.weitzlux.com), all internal development, sales/marketing efforts, public relations, and client services.
- Chief software architect.
- Increased site traffic 5,000% (to 12,000+ unique visitors/day) and original content by 2MM+ pages.
- Positioned web site as a new sales channel, netting \$MM in new revenue, 45,000% increase in leads to 1,500 new case leads/week.
- Gained first page ranking for 750,000+ search terms used to find the site including some of the most competitive (ex. *asbestos lawyer*).
- Reduced annual Internet marketing costs by \$3.6 million, or 96%, while increasing returns.
- Transformed internal development into a premier asset of the firm resulting in joint development efforts with Federal Express, 2 case studies on our use of Web Services by Microsoft, and the adoption of our proprietary document management technologies by other law firms across the country.
- Oversaw distribution of \$300 million/year in settlements to tens of thousands of clients.
- Increased overall internal system performance 500% with a 75% reduction in system downtime.
- Mentored technical team, developers, analysts, SEO (search engine optimization)

Technologies: C#. VB.NET, ASP.NET, SQL2005, SQL2000, OLTP, OLAP, JavaScript, HTML, Exchange2003, Win2003, IPSec, CISCO VPN, Citrix Presentation Server, Win2000, WinNT4, Project, PowerPoint, WebSense, eSafe, Active Directory (LDAP) Interface, XML, xHTML, Web Services, Visual Studio Team Foundation Server, VSS, VB6, Access, Microsoft Office, Cisco PIX (Firewall), Tape Library Backup, Document Management, CRM, ERP, Great Plains 7.0, Great Plains 6.0, Goldmine, ASP3.0, IIS 6.0, Windows Services, Snap NAS, LeadTools (Imaging), SharePoint, AscentCapture, Concordance, Summation, iConnect, ComponentOne for .NET, Flash, Mac OSX, Photoshop, SEO

**Case Studies (Business + Technology):**

[http://download.microsoft.com/documents/customerevidence/27250\\_Weitz\\_Lux\\_Final\\_BA.doc](http://download.microsoft.com/documents/customerevidence/27250_Weitz_Lux_Final_BA.doc)

[http://download.microsoft.com/documents/customerevidence/27249\\_Weitz\\_Lux\\_Final\\_CS.doc](http://download.microsoft.com/documents/customerevidence/27249_Weitz_Lux_Final_CS.doc)

**IMJ Group Inc.****2001-2002**

IMJ was a startup business/technology consulting think tank and solution implementation firm I started with two partners.

***Vice President, Technology / Founding Partner***

- Established corporate status for the firm and branding.
- Recruited and supervised staff.

Technologies: Linux, BSD, Java, Access, Microsoft Project, Microsoft PowerPoint, Microsoft Excel.

**Winstar Communications, Inc., Web Development Group****2000-2001**

The Winstar Web Development Group was an Internet consultancy, part of Winstar (a dot-com telecommunications firm), specializing in large Intranets and Internet sites.

***Acting Director of Technology, 2001;***  
***Senior Technology Architect, 2000-2001***

***Technical Leadership***

- Directed a team of 17 – project managers, developers, system engineers, and site designers.
- Managed full project lifecycle for six and seven digit budgets.
- Chief Software Architect.
- Designed Internet and Intranet sites, including Hosting Solutions.
- Established procedures and process, integrating what had been five company acquisitions into one team.
- Mentored development team.

***Business Development***

- Increased average project value 10,000% (from \$50,000) to \$5 million, including winning the team's first \$1 million contract.
- Liaised successfully with other Winstar divisions (Hosting, Office.com, Professional Services, and Winstar for Buildings) resulting in improved sales and project coordination.
- Led Winstar team on a \$50 million proposal for the NYC Board of Education with partners Sun, Oracle, Lucent, Microsoft and Compaq.

Technologies: ASP 3.0, SQL2000, SQL 7.0, JavaScript, HTML, Project, PowerPoint, VSS, Exchange2000, Win2000, WinNT4, WebSense, Active Directory, VB6, VB5, Visio, Microsoft Office, Flash, Photoshop, QuarkXpress.

Example web site: <http://www.nationalmssociety.org> still largely uses the same information architecture laid out over 18 years ago.

**Valinor, Inc.****1995-2000**

Valinor Inc., a wholly owned subsidiary of IKON Office Solutions, was the original Microsoft Solutions Partner/Provider & Authorized Technical Education Center in New York City, consulting for Fortune 500 firms.

***Technical Project Manager / Microsoft Certified Trainer, 1998-2000***  
(after graduating Columbia);

***Technical Project Leader / Senior Technical Architect, 1996-1998***  
(while attending Columbia full time 20+ hrs/wk);

***Senior Application Developer / Application Developer, 1995-1996***  
(while attending Columbia full time, 20+ hrs/wk)

*General Achievements*

- Increased average project value, formerly \$100,000 average, 500% to \$500,000
- Led teams of up to 10 developers

*Valinor Key Consulting Projects*

- **AIG/AIU NAD** — migration of AIG Europe's mainframe-based quotation and policy generation system to Microsoft Windows/Office workstation platforms, reducing the time to policy from days to minutes.
- **Berkery, Noyes & Co.** — a company-wide line of business application matching companies seeking to sell themselves with potential buyers.
- **The College Board** — Internet application to gather annual survey data for all United States accredited colleges and universities.
- **DLJ** — on site training.
- **Garban InterCapital** — a cash derivatives trading system utilized by 60 brokers.
- **IKON Office Solutions** — reconciled revenue reporting data from multiple disparate databases.
- **Matsushita** — workflow support ticketing system.
- **Maersk Shipping** — reconciled multiple shipping databases.
- **Merrill Lynch** — on site training on active trading floors.
- **NYC Department of Technology** — on site training.
- **Sumitomo Mitsui** — application to monitor, track and handle all transactions and interest calculations for a multi-billion per year transfer of high risk loans.
- **Sun Chemical** — on site training and follow up consulting.
- **Yankee Copyrights Management** — a large eCommerce application for a startup to create a marketplace for permissions to use and replicate copyrighted materials.

Technologies: ASP2.0, SQL7.0, SQL6.0, Sybase, Oracle, MTS, COM, JavaScript, HTML, Project, Powerpoint, VSS, Exchange5, WinNT4, WinNT3.51, Unix-Solaris, C++, MFC, VB6, VB5, VB4, VB3, Access, Visio, Goldmine, Microsoft Word, Excel

**ADJUNCT PROFESSOR****Columbia University School of Continuing Education, evening classes****2003-2005**

- Taught Database Design/Advanced SQL, and .NET Programming courses — QC7304, QC7305, QC7403.
- Developed curriculum for courses, part of SASE track certificate program.

**EDUCATION**

Bachelor of Arts, Economics/Mathematics, Magna cum Laude 3.72 GPA, Columbia University, 1999

Stuyvesant High School, 1995

**CERTIFICATIONS**

Microsoft Certified Professional, 1998

Microsoft Certified Trainer, 1998

**MAJOR LAW FIRMS**

LANIER LAW FIRM

BARON BUDD

SIMMONS HANLY CONROY

SEEGER WEISS

WEITZ & LUXENBERG

BERGER MONTAGUE

MAZIE SLATER KATZ & FREEMAN

BAILEY PEAVY BAILEY

KAISER GORNICK

SANDERS VIENER GROSSMAN

SALIM BEASLEY

CRONIN FRIED

NEBLETT, BEARD & ARSENAULT

BLASINGAME BURCH GARRARD ASHLEY

NICHOLS KASTER

PODHURST ORSECK

KIRTLAND & PACKARD

AYLSTOCK, WITKIN, KREIS & OVERHOLTZ

BEASLEY ALLEN

KAZAN, MCCLAIN, SATTERLEY & GREENWOOD

NAPOLI SHKOLNIK

**EXPERT TESTIMONY**

*Hearings*

**2015**

IN RE BENICAR (OLMESARTAN) PRODUCTS LIABILITY LITIGATION.  
UNITED STATES DISTRICT COURT NEW JERSEY, CAMDEN VICINAGE  
No. 1:15-md-02606

**2010**

IN RE: GADOLINIUM BASED CONTRAST AGENTS LITIGATION.  
UNITED STATES DISTRICT COURT NORTHERN DISTRICT OF OHIO, EASTERN DIVISION  
No. 1:08-GD-50000, MDL NO. 1909

**2007**

In re SEROQUEL PRODUCTS LIABILITY LITIGATION.  
UNITED STATES DISTRICT COURT MIDDLE DISTRICT FLORIDA, ORLANDO DIVISION  
No. 6:06-md-1769-Orl-22DAB.

*Depositions*

**2019**

RALPH GAMBLES, THOMAS MERCK, AND ELISE COMPO v. STERLING INFO SYSTEMS  
UNITED STATES DISTRICT COURT NEW YORK, SOUTHERN DISTRICT  
1:15-cv-09746-PAE

**2019**

SCOTT RILEY, MICHELLE KUNZA, VENUS COLQUITT-MONTGOMERY, JONATHON ALDRICH,  
AND KENDRA BUETTNER v. MONEYMUTUAL, LLC, SELLING SOURCE, LLC, AND  
PARTNERWEEKLY, LLC  
UNITED STATES DISTRICT COURT MINNESOTA  
No. 16-cv-4001 (DWF/LIB)

**2015**

PLAINTIFFS V. SMITHKLINE BEECHAM CORPORATION D/B/A,  
GLAXOSMITHKLINE

PHILADELPHIA COUNTY COURT OF COMMON PLEAS

Nos. 01144, 0402, 3694, 3678, 3672, 3758, 3686, 3727, 0489

**2015**

JOSEPH JAUHOLA VS. CANADIAN NATIONAL RAILWAY COMPANY,  
AND WISCONSIN CENTRAL, LTD.

UNITED STATES DISTRICT COURT MINNESOTA

No. 14-cv-1433

**2007**

In re SEROQUEL PRODUCTS LIABILITY LITIGATION.

UNITED STATES DISTRICT COURT MIDDLE DISTRICT FLORIDA, ORLANDO DIVISION

No. 6:06-md-1769-Orl-22DAB.

**SPEAKING ENGAGEMENTS**

*Chambers of Commerce*

- Greater NY Chamber of Commerce
- Forest Hills Chamber of Commerce
- Franklin Square Chamber of Commerce
- Medford Chamber of Commerce
- Rockville Centre Chamber of Commerce
- Oceanside Chamber of Commerce
- BRiSC - the Unofficial Silicon Alley Chamber of Commerce

*Small Business Administration*

- USSBA (March 2010)
- USSBA (April 2010)
- USSBA (May 2010)

*Business Groups and Organizations*

- Independent Business Women's Group
- Associated Builders & Contractors – Lower NY State
- Associated Builders & Contractors – Nassau/Suffolk
- East End Women's Network

*Charities and Non-Profits*

- Northport Rotary
- Kiwanis International

*Continuing Education*

- Columbia University – School of Continuing Education – adjunct professor
- Locust Valley Adult Education
- College of New Rochelle – Graduate Program
- Monroe College – Graduate Program